FeatureCloud:
Providing the worldwide first technological solution to ensure full patient data control

1st EC review meeting
Munich, 22nd – 23rd June 2020

Andreas Holzinger – WP4 Lead
Medical University of Graz – MUG (Austria)

Weihenstephan, 22nd – 23rd June 2020

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 - Overview, objectives and aims

Why:

Many AI approaches lack transparency, hence do not foster trust and ethical responsible machine learning. WP4 will work mainly in explainable AI and help the consortium towards the realization of robust machine learning approaches in order to make the results explainable to a human biomedical expert. In this context graphs play an essential role, particularly for causality research to answers questions of why.

How:

Create a solid overview and understanding of feature spaces (e.g. graphs), and the visualization of histopathological decision making to form the underlying knowledge base.

Foresee the development of a pathology AI workflow/application, consisting of three modules implemented as Docker containers.

Shape and compose feature spaces (e.g., graphs) for medical data which are vital in information propagation behavior, consequently the goal is to explore whether and to what extent certain modalities (e.g. graphs) are useful for privacy-aware medical data processing.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 - Overview, objectives and aims

- **Objectives and progress:**
- Objective 1: to create a solid overview on graph parallelism and to form the underlying knowledge base for all our later endeavours (100 %)
- Objective 2: to foresee the overall topology of a graph that was never seen in its entirety but only implicitly present via its distributed subgraphs and to experiment whether and to what extent such a graph can be thought of being connected in the first place; a sub goal is the exploration of link prediction via node similarity or feature-feature interaction as a necessary pre-processing step (100%)
- Currently working on the use-case of digital pathology (30 %)

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 – Milestones & Deliverables since project start

No.	Title	M	Date	Status
M24	Solid overview and understanding of state-of-the-art graph parallelism	6	Jun 19	✓
D4.1	Survey on graph parallelism	6	Jun 19	✓
D4.2	Test report on different graph types	12	Dec 19	✓

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 – Scientific advances and results since project start (Part 1 Graphs)

Graph parallelism & graph types

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

General goals & objectives

1. Explainable & privacy-aware Machine Learning
2. ... on high-dimensional data
3. ... by using graphs as intuitive & universal data structures to depict topology and relations between entities

- Graph-based / hybrid (= in addition to NNs) approaches have several advantages:

1. Form naturally on many interesting data sets (see next slide)
2. Easier to interpret (in structure as well as results) - in part because they inherently project high-dimensional data into a 2D/3D space
3. Allows for a multi-stage & distributed system where each signal conveys meaning in itself (interest update, recommendation, chemical reaction, ...)

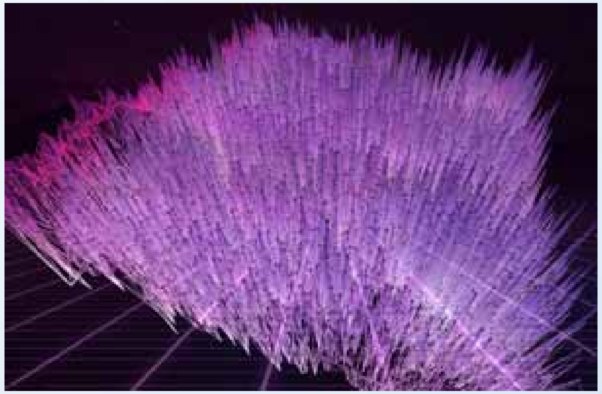
This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Graphs as universal data structures

Social network

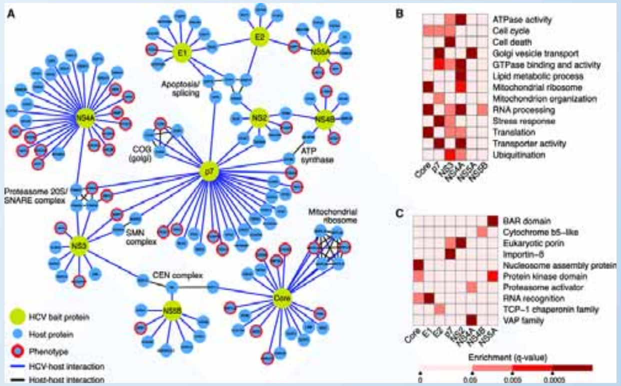
http://socialengineindia.com/

Cell image => graph



Holzinger, Andreas & Malle, Bernd & Giuliani, Nicola. (2014). On Graph Extraction from Image Data. 552-563. 10.1007/978-3-319-09891-3_50.

Protein-protein IN



Ramage, Holly & Kumar, Gagandeep & Verschueren, Erik & Johnson, Jeffrey & Dollen, John & Johnson, Tasha & Newton, Billy & Shah, Priya & Horner, Julie & Krogan, Nevan & Ott, Melanie. (2015). A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay. Molecular cell. 57. 329-340. 10.1016/j.molcel.2014.12.028.

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Learning on distributed graphs - Challenges

FeatureCloud

1. Expensive communication, since all nodes need to update a central (infrastructure of) servers according to their local model evolution
2. Systems Heterogeneity, meaning that edge devices might be of severely different storage & computational capacity
3. Statistical Heterogeneity, since each local agent might have their own objective function and subsequently differently distributed data sets (I.I.D. assumption does not hold).
4. Privacy concerns, meaning that model updates (even without transmitting the underlying data) can reveal sensitive information, while current approaches to counteract this phenomenon (e.g. adding noise via a differential privacy model) can significantly reduce systemefficiency.

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Graph convolutional networks & feature propagation

- Graphs share “hierarchical feature layers” with images, but not their spatial locality (rigid grid-pattern)
- In recent years, several methods have been established to learn on huge graphs by propagating & coalescing features amongst nodes
- Although these methods have been successful in large enterprise settings, the question is how to implement them in a distributed way using resources on the “edge” (as opposed to the backbone)
- Also, the theory behind GCN’s is not yet fully established.

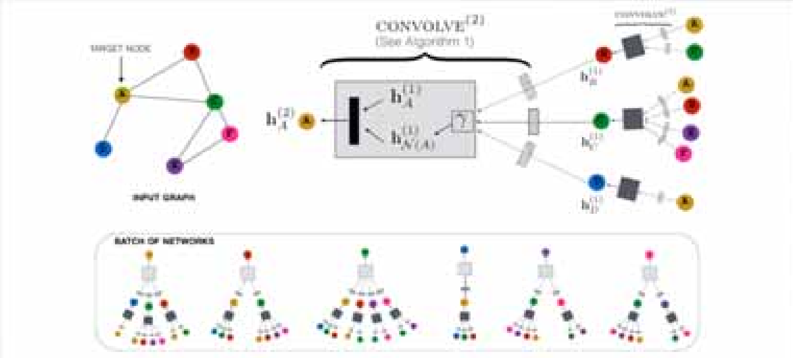


Figure 1: Overview of our model architecture using depth-2 convolutions (best viewed in color). Left: A small example input graph. Right: The 2-layer neural network that computes the embedding $h_A^{(2)}$ of node A using the previous-layer representation, $h_A^{(1)}$, of node A and that of its neighborhood $\mathcal{N}(A)$ (nodes B, C, D). (However, the notion of neighborhood is general and not all neighbors need to be included (Section 3.2).) Bottom: The neural networks that compute embeddings of each node of the input graph. While neural networks differ from node to node they all share the same set of parameters (i.e., the parameters of the CONVOLVE$^{(1)}$ and CONVOLVE$^{(2)}$ functions; Algorithm 1). Boxes with the same shading patterns share parameters; γ denotes an importance pooling function; and thin rectangular boxes denote densely-connected multi-layer neural networks.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 974–983, 2018. doi: 10.1145/3219819.3219890

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

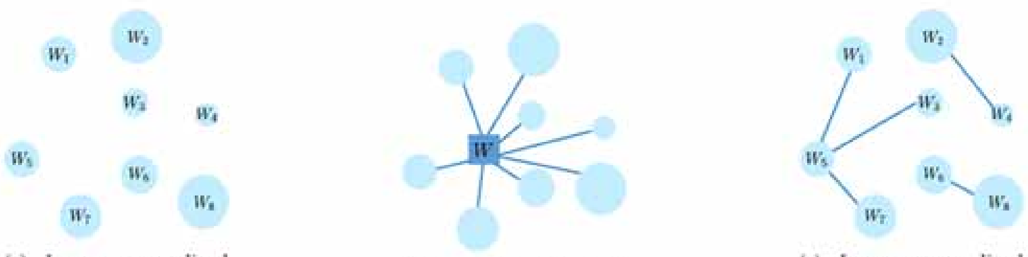
Federated Learning - traditional

- Top image: Traditionally (as introduced by Google), the goal of FL is to learn a global model from distributed data:
- A global (pre-trained model) is distributed to all client devices -> each one gets exactly the same!
- As users interact with the client, the model gets updated individually, resulting in a specialized model over time.
- In intervals, clients compute diffs & send them to the server, the sum of which are reconciled into a new global model, which again is distributed downstream.

- Bottom image: 3 possible modes of FL:
- 1) individual, 2) global, 3) learn from peers

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

https://blog.ml.cmu.edu/2019/11/12/federated-learning-challenges-methods-and-future-directions/



(a) Learn personalized models for each device; do not learn from peers.

(b) Learn a global model; learn from peers.

(c) Learn personalized models for each device; learn from peers.

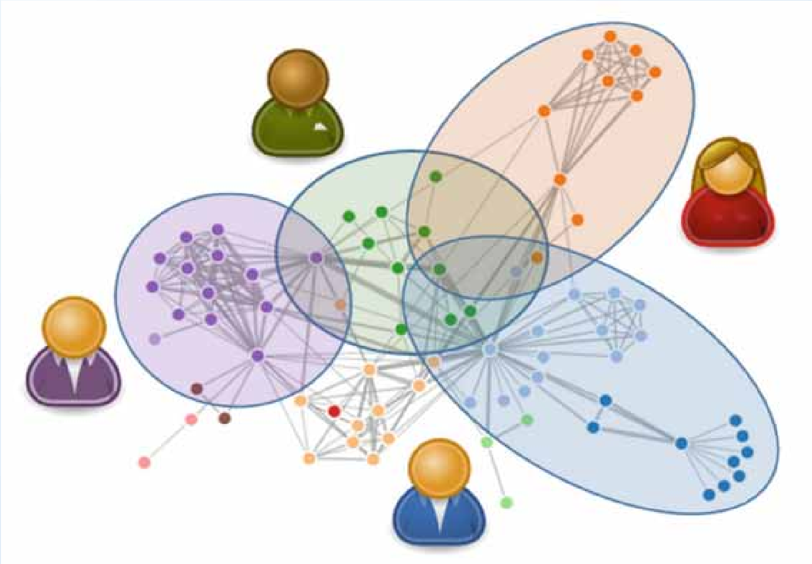
Figure 5: Different modeling approaches in federated networks. Depending on properties of the data, network, and application of interest, one may choose to (a) learn separate models for each device, (b) fit a single global model to all devices, or (c) learn related but distinct models in the network.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated Learning: Challenges, Methods, and Future Directions. pages 1–21, 2019. URL http://arxiv.org/abs/1908.07873.

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Local spheres - going beyond federations

- In human society, solutions to complex problems are usually solved by a collaboration of experts, each contributing their unique talents - NOT by smoothing out the knowledge base of the collective.
- Likewise, allowing each node to retain a local model fitting their respective data / objective function might lead to better global results even in the absence of a global model.
- This way, each local model would act as a re-usable component - and the swarm could adapt to new problems without continually re-training from the start

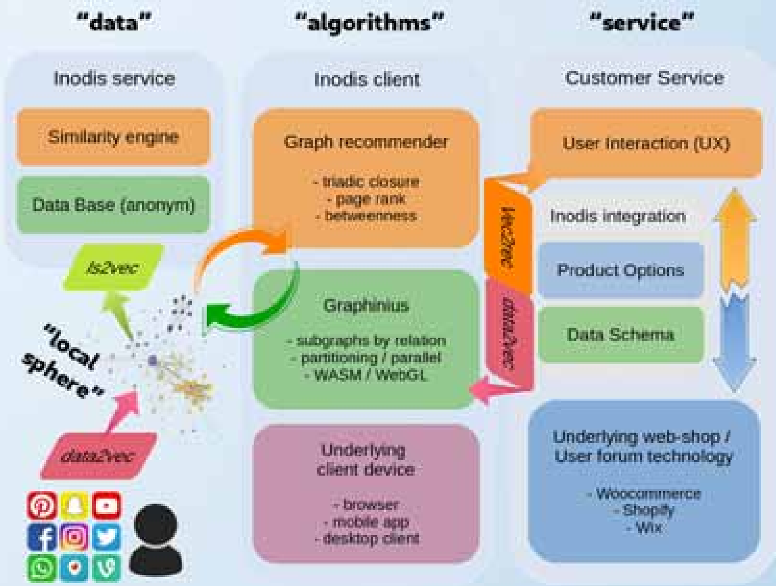


Bernd Malle, Nicola Giuliani, Peter Kieseberg, and Andreas Holzinger. The More the Merrier - Federated Learning from Local Sphere Recommendations. In Machine Learning and Knowledge Extraction, IFIP CD-MAKE, Lecture Notes in Computer Science LNCS 10410, pages 367–374. Springer, Cham, 2017. doi: 10.1007/978-3-319-66808-6 24.

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Local spheres - local / global connection

- Efficient communication among relevant local spheres is crucial
- We want to avoid “gossip-style” networks (everyone talking to everyone else)
- We propose computing a “fingerprint” per local sphere in the form of a representative feature vector (ls2vec)
- This way, a central similarity service can “connect” local spheres most likely to profit from each other’s local model / expertise.



Bernd Malle - Inodis system architecture

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 – Scientific advances and results since project start (Part 2 Use Case for Feature Cloud)

Pathology AI App ("The FeatureCloud Use-Case")

MUG is developing an AI app to aid pathologists make their diagnoses
It's aim is to fully leverage the federated approach offered by the Feature Cloud platform

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Pathology AI App

The MUG is developing an AI app to aid pathologists make their diagnoses

It’s aim is to fully leverage the federated approach offered by the Feature Cloud platform

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Pathology AI App

- The Medical University has been working on a **Pathology AI workflow** for the Feature Cloud platform
- It **currently** consists of the following **Docker** container modules:
  1. **Learner**: ran locally by each institution on local data to train a model
  2. **Evaluator**: used to test models or ensembles of models
  3. **Visualiser**: used to visualise results or interact with a final model
- The purpose of the Pathology AI App was a proof of concept.
  - We wanted to see if our requirements for an app to be used by pathologists could run on the Feature Cloud platform
  - For example, it was important that web-based apps would run and could be controlled properly by the Feature Cloud controller
- The next sections describe the app

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Learner

- The **Learner** is responsible for training a local model.
  - The **Learner** is ran by each institution locally
  - It is trained **locally**, on **local** data
  - The initial idea was simply to use ensemble learning for the final model
  - However, we do of course aim to create a properly federated approach.

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Evaluator

- The **Evaluator** is used to test local models, ensembles, or federated models on new data.
- This could work in several ways:
  - It combines models from different institutions and tests the ensemble model on a held back test set
  - It is used to evaluate one local model on a test set that is never seen locally
- Basically the idea was to be able to test the usefulness of models across institutions

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Visualiser

- The **Visualiser** is a web-based application for viewing or interacting with a trained model
- In the case of the **Pathology AI App** the Visualiser uses a trained model to perform classifications/predictions, and display some metrics
- The trained model could be a single model, or an ensemble of models or single federated model that is shared between all participants

We will now describe how the Visualiser was built

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Visualiser: GUI and Model

- The Visualiser was written in **Shiny**.
  - This is a framework for **R** that allows you to create web-based visualisations and applications quickly
- The model itself was trained using **Python** and Keras/TensorFlow
  - We used the VGG/OxfordNet 16 layer network, pre-trained on ImageNet images
  - Training was done using approximately 2,500 histo-pathology images
- Even though the model was trained in Python it is run using R and the Keras R package in the Shiny app

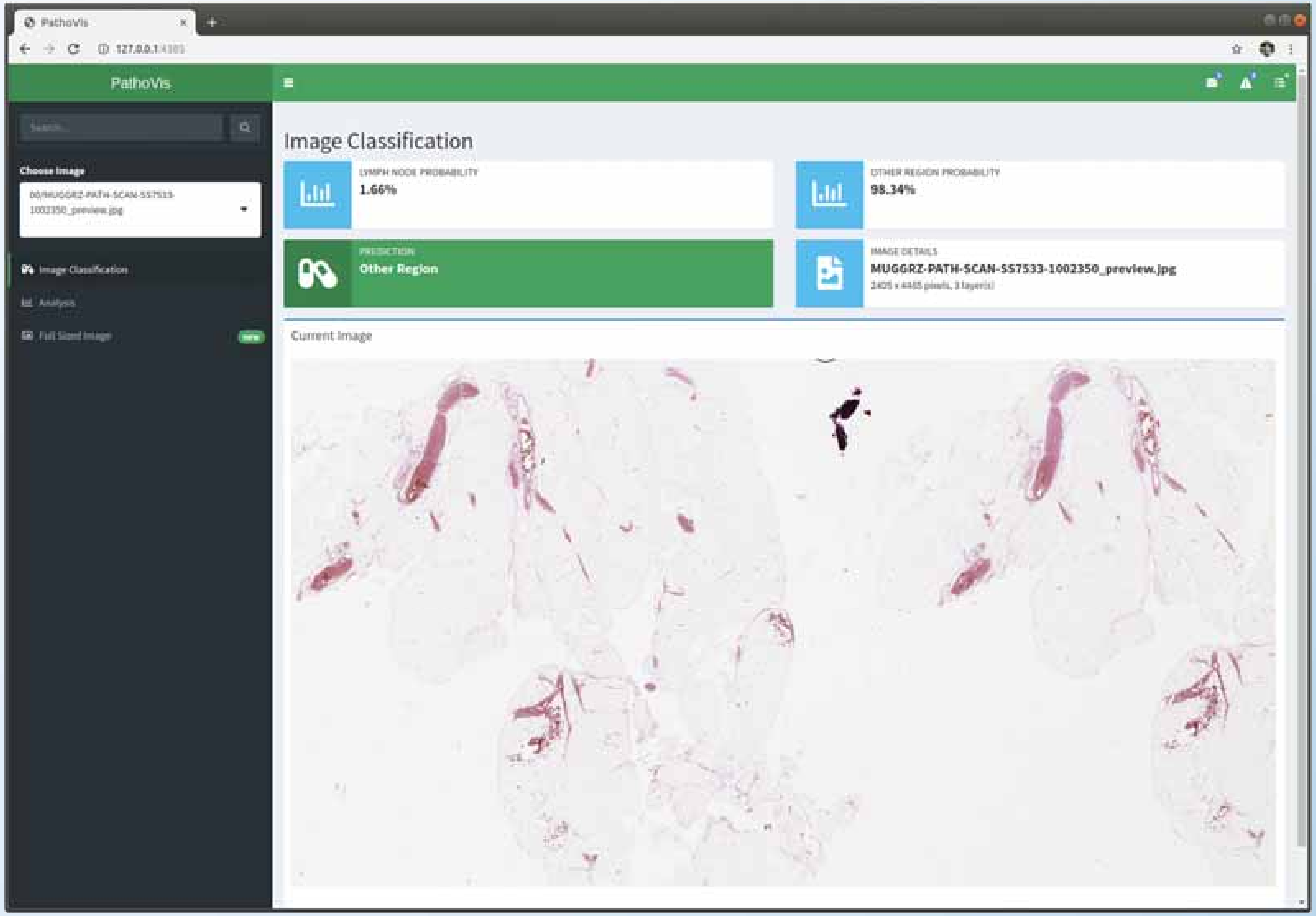
This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Visualiser: Requirements

- In order to work within the Feature Cloud platform, the Shiny application had to interface with Feature Cloud using the REST API
  - Plumber was used to write the REST layer
- We created templates to get up and running quickly with Shiny, R, Keras, and a REST API: https://gitlab.lrz.de/mdbloice/shiny-minimal-example
  - Using this minimal example, it would be easy to get a web-based app running on Feature Cloud with little effort
- It was important for the MUG that a web-based app could function on the Feature Cloud platform (more on why later)

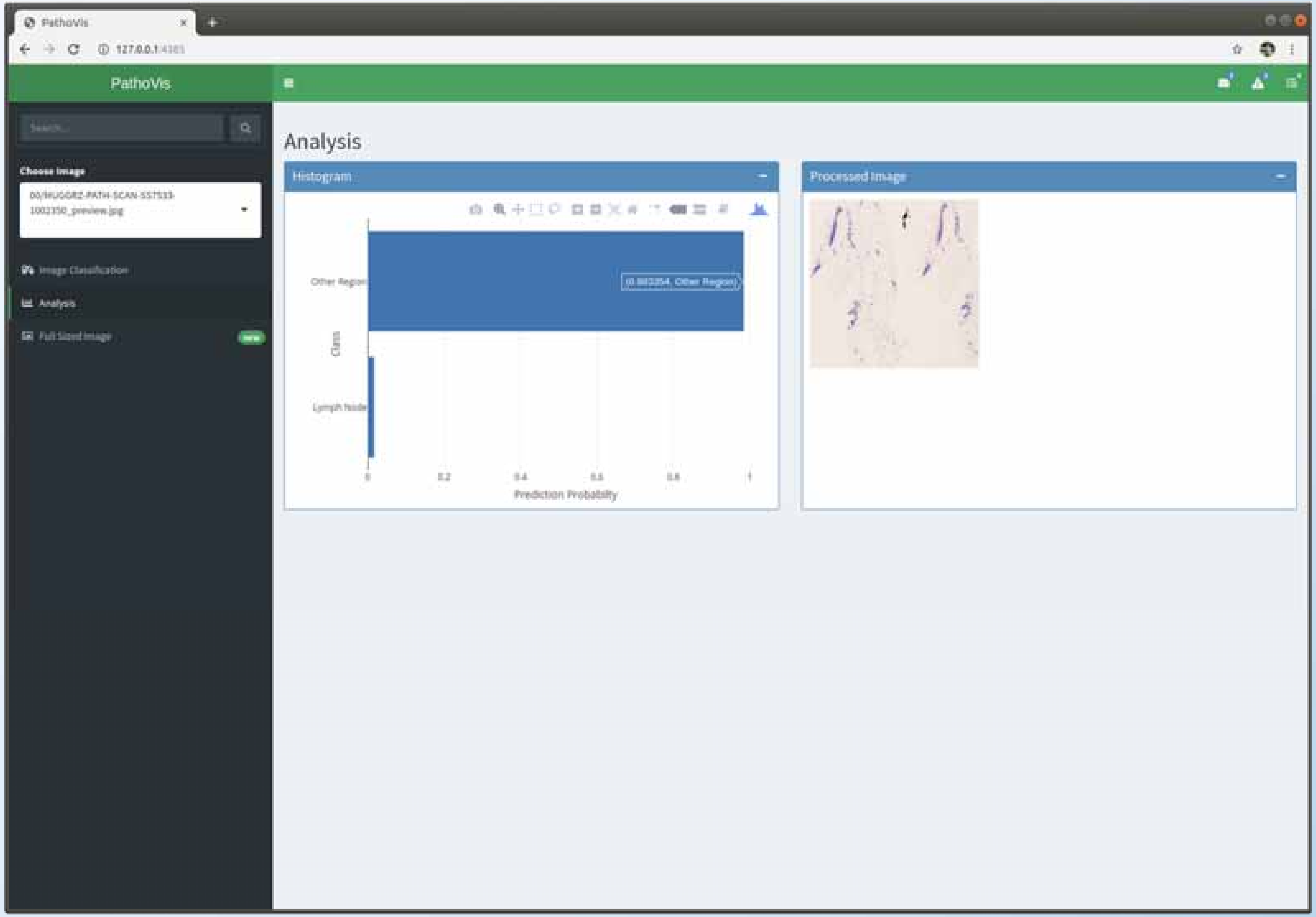
This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Visualiser



This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Visualiser



This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Pathology AI App

- The Pathology AI App was a **proof of concept**, to see if the required tools could function on the Feature Cloud platform
  - It was trained to classify if images were obtained from lymph nodes or not
  - Classification accuracy was about 95%
  - However, this is a trivial task for a pathologist
  - But, it was necessary to build a prototype system for requirements analysis
- For example, for the MUG, it is important that the app can be developed as a web app, as it is close to impossible for any other types of application to run on the hospital LAN
- Therefore, it was important that it was feasible to run Shiny applications within Feature Cloud
- **Moving forward** we require a realistic use case. For this we will concentrate on ovarian cancer whole slide image classification

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Moving Forward: Ovarian Cancer AI App

- In order to create a useful application that utilises federated learning we need a **proper use case**
- For this we will create an ovarian cancer AI app that should diagnose histo-pathology whole slide images
- **All data for this project is open**, and gathered via the NIH's Genomic Data Commons Data Portal
- This will make it easy to demonstrate the project, will be useful for testing, and we will require no ethics approval for training the algorithm, distributing it, etc.
- Also, Whole Slide Images make for **a very useful use case for federated approaches in general**…

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

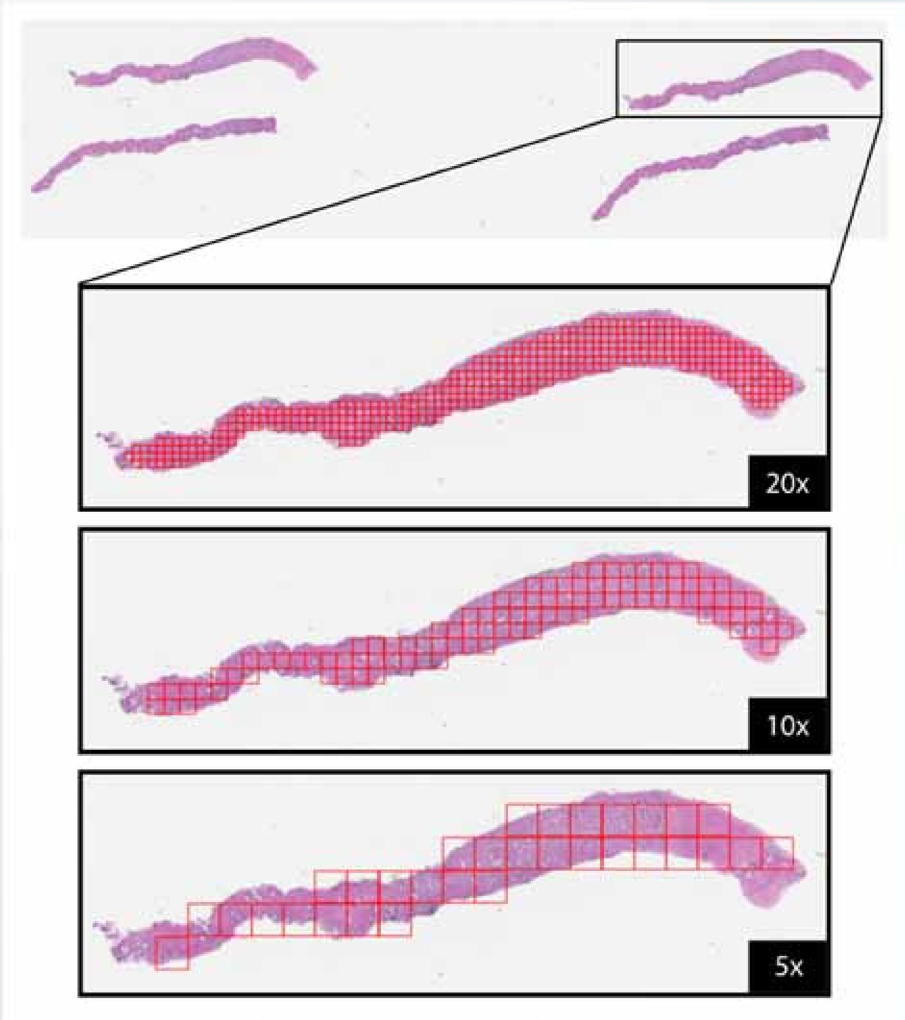
Ovarian Cancer AI App

- Because whole slide images are so large they **make a good use case for federated learning**
  - An image can be 35,000 by 65,000 pixels or larger
  - They can be many GB in size per image
- This is problematic for two reasons:
  - Rarely do institutions have the GPU memory resources to handle such large images
  - The solution to this memory problem is normally to tile the images: but then you have a computational problem in terms of time

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Ovarian Cancer AI App

- How are Whole Slide Images analysed if they are so large?
- The general approach is that images are **tiled**, so that they fit in to the memory (image)
- Several approaches use this, including *Multiple Instance Learning (Campanella G, Silva V, Fuchs T: Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology. arXiv:1805.06983, 2018)*
- However, this is still problematic for centres with little resources…



This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Ovarian Cancer AI App

- For example, in the paper mentioned above, the authors tiled their whole slide images
- This meant the images could fit in to the memory of the GPUs they used to train the algorithm
- However, they still required huge computational resources
  - They used four DGX workstations worth about $750,000 in total
  - These trained for 1 week for a set of 10,000 whole slide images
- This is because whole slides images are so large, and the amount of tiles you create per whole slide is enormous
- Therefore, we still require an approach that is feasible for the computational resources available to most groups

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

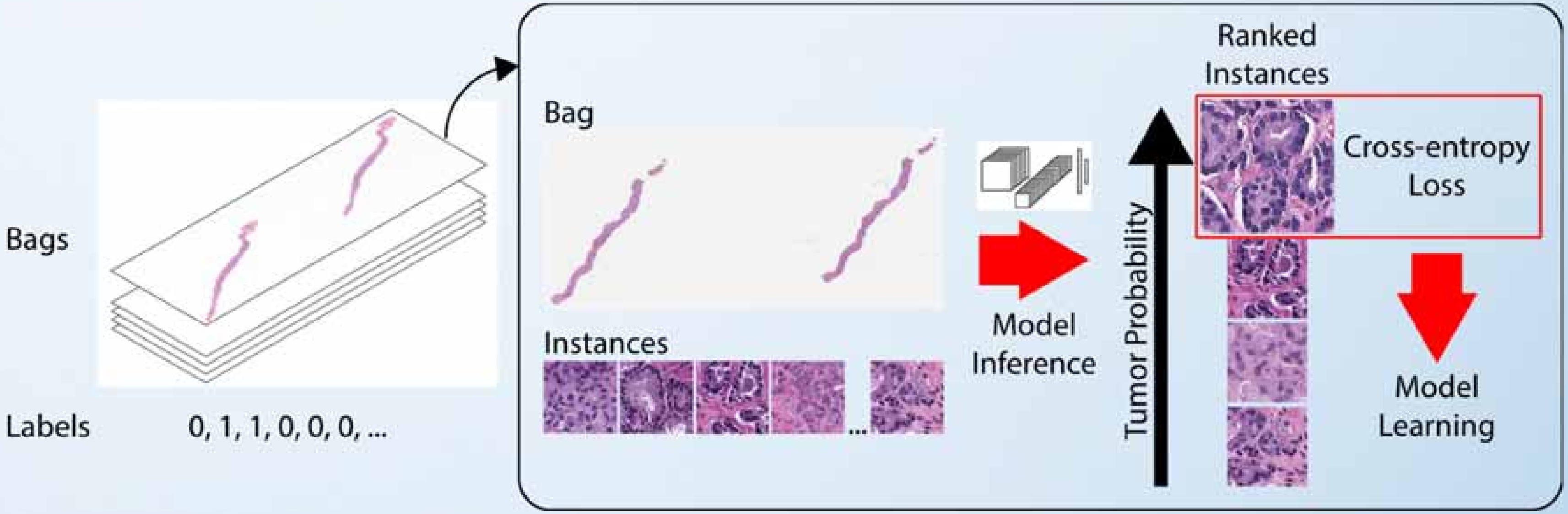
Ovarian Cancer AI App

FeatureCloud

- A federated approach could be a solution to this and makes for a good use case
- There is no reason why something like Multiple Instance Learning (or some modification of it) could not operate in federated way
- For example, in the case of Multiple Instance Learning
  - Groups of tiles from a single image could be sent to distributed institutions
  - A main model is updated
  - **or**
  - Each group trains locally on a subset of whole slide images
  - A main model is updated

This project hast received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 826078

Federated AI App



Campanella G, Silva V, Fuchs T: *Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology*. arXiv:1805.06983, 2018.

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

Summary

- Therefore, the approach that the MUG will take is to
  - reduce the memory burden (by tiling the images) and
  - reduce the computational burden (by distributing the effort)
- However, the exact approach is not yet finalised
- With such an approach we can remain competitive with institutions that have far more hardware
- It makes for a compelling use case for federated learning in a real world setting

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078

WP 4 – Backup slides

Please prepare an extensive set of back up slides to be able to respond to potential questions of the reviewers.

This project hast received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078